Recombination Rate and the Distribution of Transposable Elements in the *Drosophila melanogaster* Genome

Carène Rizzon, Gabriel Marais, Manolo Gouy, and Christian Biémont[1]

Laboratoire de Biométrie et Biologie Evolutive, Unité Mixte de Recherche Centre National de la Recherche Scientifique 5558, Université Lyon 1, Cedex, France

We analyzed the distribution of 54 families of transposable elements (TEs; transposons, LTR retrotransposons, and non-LTR retrotransposons) in the chromosomes of *Drosophila melanogaster*, using data from the sequenced genome. The density of LTR and non-LTR retrotransposons (RNA-based elements) was high in regions with low recombination rates, but there was no clear tendency to parallel the recombination rate. However, the density of transposons (DNA-based elements) was significantly negatively correlated with recombination rate. The accumulation of TEs in regions of reduced recombination rate is compatible with selection acting against TEs, as selection is expected to be weaker in regions with lower recombination. The differences in the relationship between recombination rate and TE density that exist between chromosome arms suggest that TE distribution depends on specific characteristics of the chromosomes (chromatin structure, distribution of other sequences), the TEs themselves (transposition mechanism), and the species (reproductive system, effective population size, etc.), that have differing influences on the effect of natural selection acting against the TE insertions.

Transposable elements (TEs) have been found in organisms as different as bacteria, nematodes, yeast, plants, fishes, and mammals including humans. Evidence is accumulating that they are agents of genome restructuring and, as such, appear to be a major constituent of genomes (Kidwell and Lisch 1997; Shapiro 1999; Tomilin 1999). TEs have a transposition capacity that enables them to invade the genome, leading to insertional mutations and chromosomal rearrangements. Therefore, organisms have developed various mechanisms to limit their number. However, the relative importance of the forces that control the dynamics of TEs in natural populations is still controversial (Biémont et al. 1997; Charlesworth et al. 1997; Kidwell and Evgen'ev 1999; Nuzdhin 1999). It has been proposed that containing the number of TE copies must involve either selection against rearrangements caused by ectopic recombination between TE insertions (Langley et al. 1988; Charlesworth et al. 1994, 1997; Zhang and Peterson 1999; Gray 2000) or selection against TE-induced mutations (Biémont et al. 1997). There are data consistent with both of these hypotheses. For example, in humans, ectopic exchange between Alu sequences seems to be more important in producing deleterious mutations (0.2%–0.3% of diseases) than insertional mutagenicity (0.1%; Roy et al. 1999), whereas ~50% of mutations in *Drosophila melanogaster* are attributable to TE insertions (Finnegan 1992). If the ectopic exchange in a region is proportional to the meiotic exchange in that region (Langley et al. 1988; Petes and Hill 1988; Montgomery et al. 1991; Goldman and Lichten 1996, 2000), then the number of TE insertions should be negatively correlated with the recombination rate. The same is true for selection against insertional mutations, because selection is weaker in regions of low recombination (Hill and Robertson 1966, Charlesworth et al. 1993; Kliman and Hey 1993). Hence, the density of FL L1 elements has been found to be negatively correlated with recombination rate in humans (Boissinot et al. 2001), suggesting that purifying selection against TEs is occurring. However, the distribution of TE insertion sites over the chromosomes of *D. melanogaster* shows no evident relationship between the frequency of recombination and TE density (Hoogland and Biémont 1996), although it is well known that TEs accumulate in heterochromatic regions such as the chromocenter and the bases of chromosomes, which are characterized as sites where there is little or low recombination (Charlesworth et al. 1992a,b). The TEs of *Drosophila* also seem to be more abundant on chromosome 4 and within some inversions, both of which have a low recombination rates (Montgomery et al. 1987; Langley et al. 1988; Eanes et al. 1992; Sniegowski and Charlesworth 1994). In plants, many elements are located in clusters at the paracentric heterochromatin (Brandes et al. 1997), the *copia*-like elements are concentrated in the centromeric regions (Heslop-Harrison et al. 1997), and the regions flanking the centromeres are densely populated by TEs (Copenhaver et al. 1999). It is becoming increasingly evident, however, that TEs are major constituents of the centromeric regions in *Drosophila* (Pimpinelli et al. 1995; Zuckerkandl and Hennig 1995; Pardue et al. 1996; Eissenberg and Hilliker 2000), and that this is not merely the result of passive accumulation in such regions caused by the absence of strong forces tending to eliminate them.

A recent study in *Caenorhabditis elegans* produces even more puzzling results, with a recombination rate that was found positively correlated with the amount of transposons (DNA-based elements), but not with the amount of LTR and non-LTR retrotransposons (RNA-based elements; Duret et al. 2000). The importance of the forces regulating TE distribution may thus vary according to the genome, indicating the need for comparative studies (Kidwell and Evgen'ev 1999). The difference between the findings from studies of *C. elegans* and *Drosophila* may also be attributable to the type of data used,

[1]Corresponding author.
E-MAIL biemont@biomserv.univ-lyon1.fr; FAX 33 4 78 89 27 19.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.210802. Article published online before print in February 2002.

12:400–407 ©2002 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/01 $5.00; www.genome.org

population data in the case of *Drosophila*, data from the genome of a single individual in that of *C. elegans*. Because we now possess the sequence of the entire *D. melanogaster* genome, we analyzed the distribution of its TEs in relation to the recombination rate. LTR and non-LTR retrotransposons tend to accumulate in regions of low recombination, but with no clear tendency to parallel the recombination rate along the chromosomes. There is, however, a negative correlation between the recombination rate and the density of transposons.

RESULTS

Table 1 shows the reference used for each element and the number of sequences retrieved from the *D. melanogaster* genome. Among the 54 TE families, 10 were transposons, 28 LTR retrotransposons, and 16 non-LTR retrotransposons. We thus collected 1007 insertions, 185 transposons (DNA-based elements), 572 LTR retrotransposons, and 250 non-LTR retrotransposons (RNA-based elements). No copies of *P* and *HeT-A* elements were identified in the genome. Figure 1 reveals that TEs accumulate mainly in pericentromeric regions and in chromosome 4, but not in the telomeric regions, which is consistent with what is usually observed (Charlesworth et al. 1992a,b).

TE Density According to Recombination Rate

Because the accumulation of TEs in *Drosophila* heterochromatin is likely to be a consequence of the peculiar properties of the heterochromatic material (Pimpinelli et al. 1995), it is not possible to establish a direct relationship between the accumulation of TEs in heterochromatin and the recombination rate (Garcia Guerreiro and Fontdevila 2001). As the pericentromeric regions and chromosome 4 are heterochromatic regions, the relationship between TE density and recombination rate was studied both with and without these regions to avoid statistical bias. Moreover, as telomeric regions have peculiar evolutionary dynamics (Marais et al. 2001), we also performed the analyses with and without these regions.

Figure 2 shows the relationship between the mean density of TEs and the recombination rate for each TE type. The density of LTR retrotransposons (Fig. 2a) was high for low values of recombination, but was homogeneously distributed afterward, and the density of non-LTR retrotransposons did not seem to follow any clear tendency (Fig. 2b). However, the Spearman rank correlations were significant ($\rho = -0.13$, $P < 0.01$ for LTR retrotransposons, $\rho = -0.15$, $P < 0.01$ for non-LTR retrotransposons), and the χ^2 values, calculated for four classes of recombination each containing 25% of the values, were also significant ($\chi^2 = 113.12$, $P < 0.0001$ for LTR retrotransposons, $\chi^2 = 52.46$, $P < 0.001$ for non-LTR retrotransposons), with accumulations of insertions in the lower recombination rate class. We did the calculation again, this time without the pericentromeric and telomeric regions and chromosome 4. The Spearman rank correlations were no longer significant ($\rho = -0.02$, $P > 0.05$ for LTR retrotransposons, $\rho = 0.08$, $P > 0.05$ for non-LTR retrotransposons), although the chi^2 statistics were still highly significant for LTR retrotransposons ($\chi^2 = 34.42$, $P < 0.0001$) and to a lesser degree, significant for non-LTR retrotransposons ($\chi^2 = 8.94$, $P = 0.03$), as a result of accumulations of insertions in the lowest recombination class value. Therefore, we conclude that the densities of LTR and non-LTR retrotransposons do not parallel the recombination rate along the chromosomes, although these TEs do tend to accumulate in regions with low recombination rates.

Table 1. Number of Retrieved Copies of the *D. melanogaster* Transposable Elements Analyzed

Element	Reference	Number of copies
LTR retrotransposons		**572**
17.6	X01472[a]	18
297	X03431[a]	69
412	X04132[a]	32
1731	X07656[a]	15
Aurora	X70361[a]	5
Bel	U23420[a]	11
Blastopia	Z27119[a]	19
Blood	X04671[a]	26
Burdock	U89994[a]	15
Circe	X98424[a]	1
Copia	X04456[a]	32
Cruiser	AF364550[a]	24
Gate	AJ010298[a]	1
Gypsy	M12927[a]	8
HMS-beagle	J01078[a]	17
Idefix	AJ009736[a]	9
Mdg1	X59545[a]	26
Mdg3	X95908[a]	14
Micropia	X14037[a]	2
Nomad/yoyo	AF039416[a]	23
Pilgrim	Costas et al. 2001	5
Roo/B104	U11691[a]	123
Springer	D17529[a]	10
Stalker	EBI[b]	29
Tinker	EBI[b]	9
Tirant	X93507[a]	22
Transpac	AF222049[a]	6
Zam	AJ000387[a]	1
Non-LTR retrotransposons		**250**
BS	X77571[a]	10
Doc	X17551[a]	55
F	M17214[a]	42
G	X06950[a]	7
Helena	AF012030[a]	3
HeT-A	U06920[a]	0
I	M14954[a]	20
Jockey	M22874[a]	38
Pilger	AJ278684[a]	1
R1Dm	X51968[a]	4
R2Dm	X15707[a]	5
Tart	U02279[a]	1
WaldoA	EBI[b]	27
WaldoB	EBI[b]	13
X	AF237761[a]	17
You	AJ302712[a]	7
Transposons		**185**
Bari1	X67681[a]	5
FB	X15469[a]	1
HB1	X01748[a]	30
Hobo	X04705[a]	20
Hoppel	M55078[a]	73
Hopper	X80025[a]	11
P	X06779[a]	0
Pogo	X859837[a]	16
S	U33463[a]	28
Vivi	S78304[a]	1

[a]Accession number in Genebank.
[b]Reference in EBI (the European Bioinformatics Institute).

There was a significant negative correlation between the density of transposons and the recombination rate (Spearman rank correlation coefficient, $\rho = -0.29$, $P < 0.0001$; Fig. 2c), which remained significant after eliminating the pericentromeric, telomeric, and chromosome 4 regions ($\rho = -0.19$,

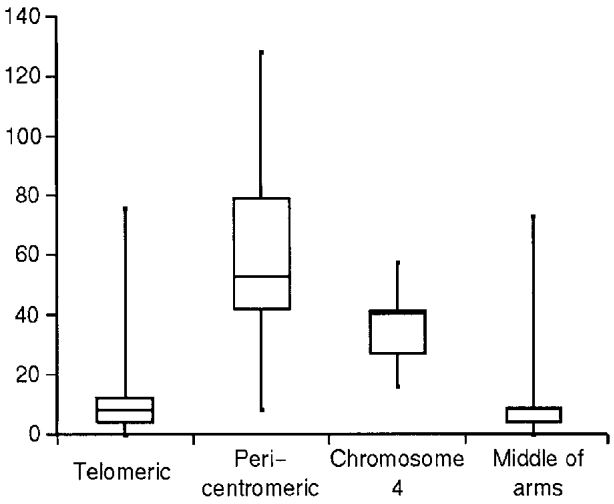

Figure 1 Box diagram of TE densities according to specific chromosomic regions, showing the minimal, quartile 1, median, quartile 3, and maximal values of the TE densities found in 0.25-Mb genome fragments.

$P < 0.0001$). The chi^2 tests on transposon density over the four recombination classes defined above were highly significant ($\chi^2 = 197.84$, $P < 0.0001$) with an accumulation of transposons in the lowest recombination rate class when all of the genomic regions were taken into consideration. The χ^2 remained highly significant after eliminating the telomeric and pericentromeric regions and chromosome 4 ($\chi^2 = 71.78$, $P < 0.0001$). We thus conclude that the density of transposons decreases when the recombination rate increases.

We performed a detailed analysis for the LTR retrotransposons on individual chromosomes. Such analysis was not done for transposons or non-LTR retrotransposons because of the lack of data for a reliable statistical analysis. The telomeric, pericentromeric, and chromosome 4 regions were not included in the analysis. No significant linear correlation between the recombination rate and the density of LTR retrotransposons was found for any of the chromosome arms (Spearman rank correlation, $\rho = -0.13$, $P > 0.05$ for 2L; $\rho = -0.12$, $P > 0.05$ for 2R; $\rho = -0.16$, $P > 0.05$ for 3L; $\rho = -0.09$, $P > 0.05$ for 3R), although all values were negative. These negative correlation values were due to an accumulation of LTR retrotransposon copies in regions of low recombination, which was significant for the 2L arm (Fig. 3a: $\chi^2 = 43.33$, $P < 0.0001$) and the 3R arm (Fig. 3d: $\chi^2 = 14.56$, $P < 0.01$), but not for the 2R or 3L arms. The X chromosome gave different results. Because of the narrow range of recombination values in the middle portion of this chromosome, the distribution of LTR retrotransposon insertions was analyzed along the entire chromosome by use of nonparametric statistics. The X arm was split into 5000-bp fragments, which were coded 1 if they contained at least one TE insertion and 0 if no insertion was detected. This allowed us to calculate the Variance of Ranks, VR, which detects aggregation in the middle of the sequence of 0 and 1 (low variance value) or at its ends (high variance value), and the Multiple Pool, MP, which detects aggregation at various regions in the sequence (see Aulard et al. 1995, for details of these statistics). These tests (VR = 3.5, $P < 0.001$; MP = 3.73, $P < 0.001$) showed that copies of LTR retrotransposons were at least grouped near the X centromere. A one-sample Kolmogorov-Smirnov test, which detects aggregation in a given region, revealed a significant group of LTR retrotransposon copies around 3.33 Mb, outside but not far from the telomeric region (KS = 0.15, $P < 0.01$). Therefore, there was no tendency for LTR retrotransposons to accumulate in the X telomere despite its low recombination rate.

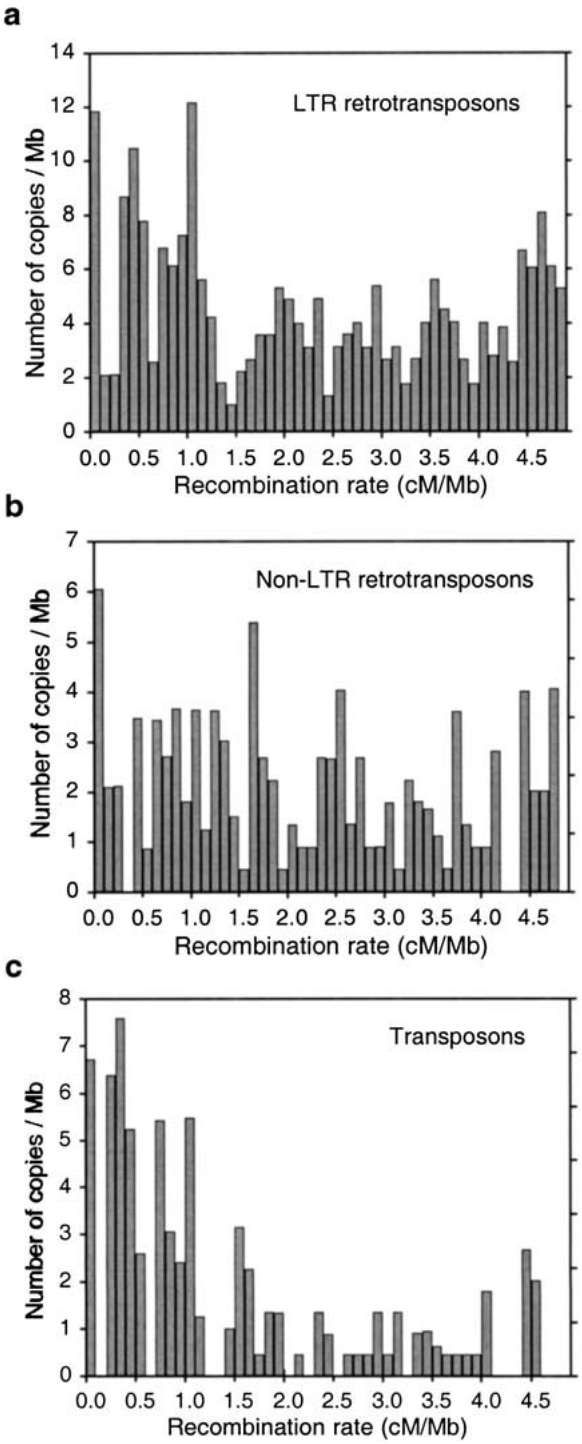

Figure 2 Density of LTR retrotransposons (*a*), non-LTR retrotransposons (*b*) and transposons (*c*) according to recombination rate of the genome of *Drosophila melanogaster*.

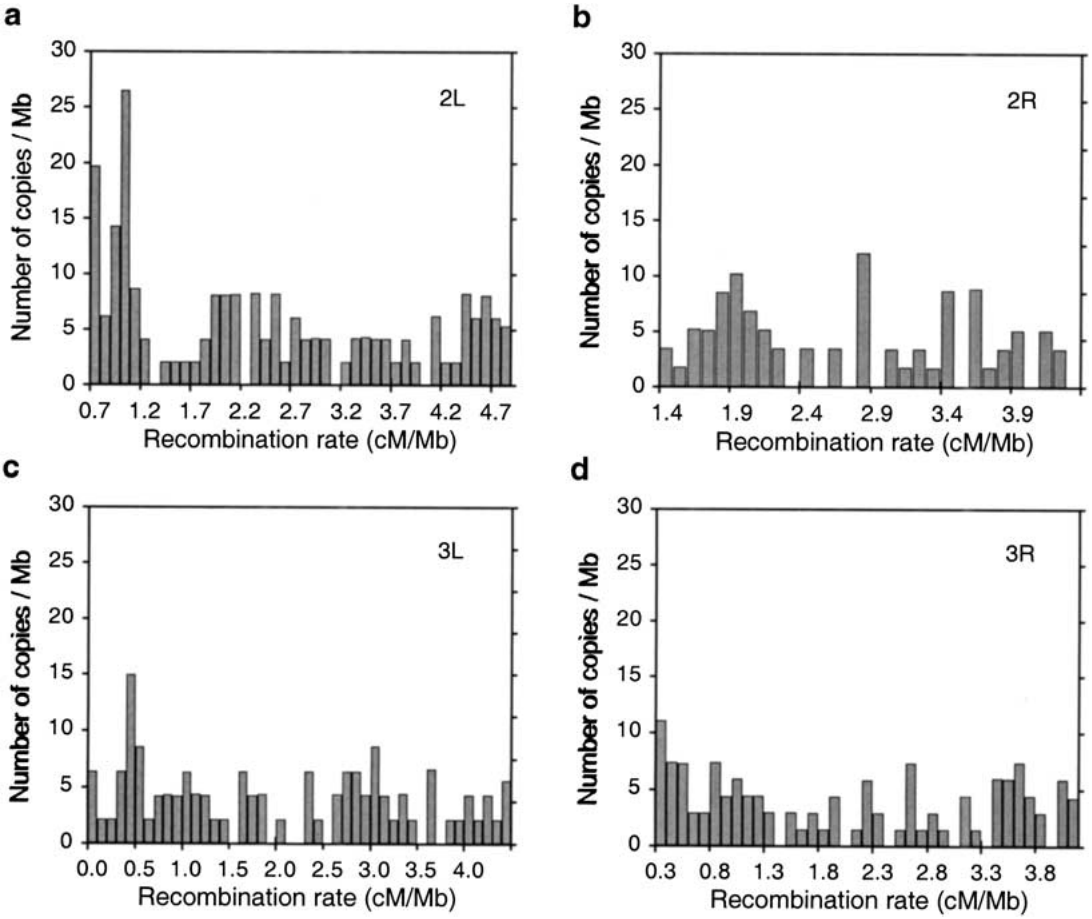

Figure 3 Density of LTR retrotransposons according to the recombination rate in the chromosome arms 2L (*a*), 2R (*b*), 3L (*c*), and 3R (*d*) in the genome of *D. melanogaster*.

To make it easier to compare our data with that from the population analysis (Hoogland and Biémont 1996), we calculated the Spearman rank correlation between recombination rates and the mean density found in each recombination rate class of the elements *roo*/*B104*, *mdg3*, *mdg1*, *copia*, *412*, *297*, *I*, and *hobo*, for which we had reliable data from both studies. Once again, the pericentromeric, telomeric, and chromosome 4 regions were excluded. The ρ values were not significant, and fell within the range of values of the population study (*roo*/*B104*, $\rho = -0.14$, $P > 0.05$; *mdg3*, $\rho = -0.25$, $P > 0.05$; *mdg1*, $\rho = 0.16$, $P > 0.05$; *copia*, $\rho = 0.25$, $P > 0.05$; *412*, $\rho = -0.10$, $P > 0.05$; *297*, $\rho = -0.21$, $P > 0.05$; *I*, $\rho = -0.04$, $P > 0.05$; *hobo*, $\rho = -0.14$, $P > 0.05$), suggesting that there was no great difference between the two approaches.

Gene Density According to Recombination Rate

No significant correlation was detected between gene density and recombination rate (Spearman rank correlation, $\rho = -0.09$, $P > 0.05$), suggesting that the relationship between TE density and recombination rate was not biased by the amount of genes.

TE Density on Autosomes and on the X Chromosome

We compared the TE density on autosomes and the X chromosome for each class of TEs and for all of the TEs taken globally. When all regions were considered, the chi^2 statistics were significant for transposons ($\chi^2 = 8.40$, $P < 0.01$) and non-LTR retrotransposons ($\chi^2 = 5.48$, $P = 0.02$), both of which showed a deficit in the number of copies on the X chromosome. χ^2 was not significant for LTR retrotransposons ($\chi^2 = 3.69$, $P > 0.05$) or for all of the TEs pooled ($\chi^2 = 0.92$, $P > 0.05$). Eliminating the pericentromeric, telomeric, and chromosome 4 regions rendered the tests nonsignificant ($\chi^2 = 1.53$, $P > 0.05$ for non-LTR retrotransposons; $\chi^2 = 1.63$, $P > 0.05$ for transposons; $\chi^2 = 1.58$, $P > 0.05$ for all TEs pooled) except for LTR retrotransposons ($\chi^2 = 9.41$, $P = 0.002$), which showed an accumulation of copies on the X chromosome.

DISCUSSION

As in all previous studies, our study confirms the accumulation of TEs in centromeric and pericentromeric regions, with a low TE density in subtelomeric regions. However, after eliminating these specific regions, the densities of LTR and non-LTR retrotransposons appeared to be high in regions with low recombination rates, but to have no direct, linear relationship with the recombination rate along the chromosome arms. In contrast, a negative correlation was found between transposon density and recombination rate. The accumulation of TEs in regions of reduced recombination rate is compatible both with selection acting against deleterious mu-

tations caused by TE insertions, and with selection acting against chromosomic rearrangements caused by ectopic recombination between TE copies, because, in both cases, selection is expected to be weaker in regions of reduced recombination. The nonlinearity in the relationship between recombination and LTR and non-LTR retrotransposon densities might result from nonlinearity in Hill-Robertson effects and ectopic exchanges with meiotic recombination. However, the observations that the transposon density linearly decreased with recombination rate and that the accumulation of LTR retrotransposons was statistically significant only for a few chromosome arms suggest that selection is not sufficient to explain the distribution of TEs along chromosomes in the *D. melanogaster* genome.

An absence of linear correlation between retrotransposon density and recombination rate when centromeric and pericentromeric regions were eliminated from the calculation is congruent with previous population studies (Hoogland and Biémont 1996), in which, however, no negative correlation was detected between transposon density and recombination rate. A major difference between the two studies is that the population approach involved only two transposons, *P* and *hobo*. The present study relied on 10 transposons, without the *P* element, which was not detected in the sequenced genome. The significant negative correlation observed between transposons and recombination rate, strongly supports the hypothesis that selection acts against these TEs. But if this is so, then why is the relationship between transposon density and recombination negative in *Drosophila* and positive in the nematode (Duret et al. 2000)? Differences in meiotic pairing and recombination mechanisms could account for the contrasting relationships between TEs and recombination rate in these two species. But *Drosophila* and *C. elegans* use the same recombination-independent mechanisms to align homologs (McKim et al. 1998), and ectopic recombination is several orders of magnitude less frequent than allelic recombination in both organisms (Virgin and Bailey 1998). Yeast, however, initiates homolog colocalization and alignment by homology-dependent DNA–DNA interactions (Kleckner and Weiner 1993), and shows only small, but significant, differences between ectopic and allelic recombination frequencies (Kupiec and Petes 1988; Goldman and Litchen 1996). TEs are not randomly distributed in yeast, but are mainly located in genes transcribed by RNA polymerase III, such as tRNA genes (Kim et al. 1998). It is, however, difficult to explain why the nonuniform homolog pairing in *Saccharomyes cerevisiae* (Kleckner and Weiner 1993), as opposed to the close homolog alignment in *Drosophila* and *C. elegans*, could account for the differences between TE distribution in these two species and in the yeast genome.

Could a difference in breeding system account for the reverse correlations between transposon density and recombination rates in *Drosophila* and the nematode? If the deleterious effects of TEs are mostly recessive, then selection against TEs should be most effective in populations with high levels of homozygosity (Wright and Schoen 1999). In contrast, because ectopic exchanges occur preferentially between heterozygous TE insertions (Montgomery et al. 1991, Charlesworth and Charlesworth 1995), according to this selective model, selection should be most effective in out-crossing populations. Because the *C. elegans* breeding system is presumed to be mostly inbreeding (Baird et al. 1992), unlike that of *Drosophila*, the effects of selection can be expected to differ in these two species. Moreover, because self-fertilization can theoretically be expected to reduce the recombination rate (Morgan 2001), studying the relationship between recombination rate and TE density would not easily detect selection against TE-induced mutations in *C. elegans*. However, the present sequenced *Drosophila* genome is derived from a laboratory strain that is undoubtedly homozygous, whereas the genomes of individuals from natural populations are likely to be highly heterozygous. Therefore, homozygosity may have interfered with the mechanisms controlling the TE copy number in this specific genome during its long history in the laboratory, during which the loss or mobilization of specific TEs cannot be excluded (Biémont et al. 1987).

The contrasting relationships between recombination rate and transposon density in the nematode and *Drosophila* could be attributable to the very large population size in *Drosophila*. A large population means that selection against TE insertions are likely to outweigh drift (Charlesworth and Charlesworth 1995), so that selection can interfere with recombination, leading to fewer transposon insertions in regions of high recombination in which selection is strongest. If selection is a weaker force in the nematode, then it should be possible to detect the preference for transposons to become inserted in regions of high recombination rate, as a result of their mechanisms of transposition in this species (Duret et al. 2000). Although it is difficult to compare effective population sizes between two species, the specific reproductive system of the nematode, which is assumed to be mostly inbreeding (Baird et al. 1992), may have an effect similar to that of a smaller population. This means that the interaction between selection, effective population size, and recombination is of great importance in the structuring of genomes.

The presence of high densities of TEs in regions of low recombination and the significant negative correlation between transposons and recombination rate suggest that selection may act against the insertion of TEs. According to this hypothesis, fewer TE insertions can be expected on the X chromosome than on the autosomes. Because the males are hemizygous in *Drosophila*, deleterious TE insertions on the X should be selected against to a greater extent than insertions on the autosomes (Montgomery et al. 1987; Langley et al. 1988; Charlesworth et al. 1994). We did not detect any reduction in TE density on the X chromosome compared with the autosomes for the TEs of the sequenced genome. This observation and the fact that differences in TE amount between the autosomes and the X chromosomes have been observed for some elements, but not for others, in studies of populations of *Drosophila* (Montgomery et al. 1987; Biémont 1992; Charlesworth et al. 1992a,b; 1994; Biémont et al. 1997), suggest that selection against the insertional effects of TEs is not the main force controlling TE copy number. This is also consistent with the data on *C. elegans*, which shows no evidence that there are fewer TE insertions on the X chromosome than the autosomes (Duret et al. 2000). We must, however, consider the possibility that the TEs may have only been mobilized recently within the genome of the *Drosophila* stock used for sequencing, and that selection had not yet reduced the TE copy number on the X chromosome.

We cannot rule out the possibility that the distribution of TEs is not directly associated with recombination, but depends on other factors that could themselves be associated, possibly fortuitously, with recombination, and so depends on the individual genome considered and the species. This is illustrated by the observations that LTR retrotransposons aggregated in a small region of the X chromosome, but were

homogeneously distributed in regions of middle and high recombination rates, and accumulated in regions of low recombination other than pericentromeric regions only in the 2L and 3R chromosome arms.

The distribution of target sites for TE insertions could vary with the DNA base composition (Sharp and Matassi 1994), and this would account partly for the distribution of TEs along the chromosomes. In the *C. elegans* and *Drosophila*, the G+C content is positively correlated with the recombination rate, both in noncoding regions and in synonymous positions of codons (Marais et al. 2001). This might lead to a link between the distribution of target sites for TE insertions and the recombination rate. Many TEs seem to be inserted in AT-rich, late-replicating, DNA regions (Le et al. 2000), and their target insertion sites are often a succession of A and T. For example, the human *L1* elements show target specificity for TTTTAA, which leads to a linear negative relationship between *L1* density and GC richness. This has also been shown for LTR retrotransposons *1731*, *17.6* in *D. melanogaster*, *TRIP* in sea urchin, *Mag* of *Bombyx mori* (Springer et al. 1995), and certain retroviruses (Bernardi et al. 1985), but it has also been shown that TEs are globally AT-rich (Sharp and Matassi 1994; Lerat et al. 2000, 2002), so that insertions of numerous TE copies in a region leads to a low GC value. On the other hand, TEs could accumulate in low gene density regions, as reported for the *Arabidopsis* genome (The *Arabidopsis* genome initiative 2000), and could be associated with low GC content (Kumar and Bennetzen 1999; for review, see Lin et al. 1999; Adams et al. 2000; Jabbari and Bernardi 2000), and reduced recombination rate (for review, see Kliman and Hey 1993; Charlesworth 1994; Fullerton et al. 2001; Marais et al. 2001). In the present study, however, we did not detect any relationship between gene density and recombination rate. Moreover, the GC-rich SINE elements are located in GC-rich regions (Korenberg and Rykowski 1988; Boyle et al. 1990; Jurka 1997), and the region around *P* insertion sites in *Drosophila* are GC-rich (Liao et al. 2000). In humans, the distributions of young and old copies of Alu elements have been found to be different (Smit 1999), suggesting that Alus integrate randomly but are preferentially fixed in GC-rich DNA as the result of some force of selection. These data do not allow us to conclude that there is any specific or general relationship between base composition and TE distribution.

TEs may insert preferentially in regions in which other sequences are already inserted, making the correlation between TE density and recombination rate merely fortuitous and variable from genome to genome and species to species, especially if such sequences could be recombinogenic, as postulated for the CeRep sequences of *C. elegans* (Cangiano and La Volpe 1993; Barnes et al. 1995). If the specific sequences in which the TEs are inserted vary according to the TE family, then the association between recombination rate and TE density will vary with the TE considered. For example, microsatellites accumulate preferentially in genome regions in which recombination is infrequent (Charlesworth et al. 1994). However, the density of microsatellites is not influenced by the recombination rate in *D. melanogaster* (Bachtrog et al. 1999) and, there is no strong evidence for the overall insertion of different TEs in specific sequences, such as micro and minisatellites. We have evidence, however, that some of the microsatellite sequences may, in fact, result from the TEs themselves (Nadir et al. 1996; Jarne et al. 1998; Toth et al. 2000), although a high density of microsatellites does not always coincide with a high density of TEs, especially in *Arabidopsis thaliana* (Schlötterer 2000), even though retrotransposons have been found near microsatellites in barley (Kalendar et al. 1999).

These findings suggest that various features (local genomic composition and structure, chromatin conformation, DNA nick repair, number of DNA replications, effective population size, reproductive system, and history of the host) could variously influence and even blur the impact of natural selec tion acting against the TE insertions along the chromosome.

METHODS

Sequence Data and Locations of Transposable Elements

The sequences of the *D. melanogaster* chromosome arms X, 2L, 2R, 3L, 3R, and 4 were retrieved from the unannotated version 1 of the genome (Adams et al. 2000; BDGP 2000). The retrieved data (all tracks of N omitted) thus totaled 114.5 Mb, corresponding to 64% of the whole genome sequence (the actual sequence represented 95% of the total euchromatin). Therefore, all of the TE sequences collected were from the euchromatic part of the genome, because most of the heterochromatin, including the Y chromosome, was not sequenced. Heterochromatin is composed of many transposable elements, mostly in an inactive state and in the form of defective sequences. Hence, the TE sequences studied here represent only a fraction of all the TE sequences in the *Drosophila* genome. Our analysis is thus pertinent for comparison with data on TE chromosomal locations obtained from population analyses in which the in situ hybridization technique used gives information on the copy number of the TEs that are inserted in the euchromatin, along polytene chromosomes.

A bank of reference sequences for TE families was constituted with sequences retrieved from the FlyBase database (FlyBase 1998; http://flybase.bio.indiana.edu/). When the sequences from this database corresponded to incomplete copies of LTR retrotransposons, we searched for full-length copies with the `BLAST` program in the high-quality genomic clone sequences from BDGP/EDGP (http://www.fruitfly.org). The list was completed with sequences retrieved from EBI (The European Bioinformatics Institute 2001; http://www.ebi.ac.uk/index.html, sequence set: ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_sets) and with the *pilgrim* sequence (Costas et al. 2001). The distributions of TE insertions in the whole shotgunned genome sequence were then analyzed by use of the above reference sequences and by the program `RepeatMasker` (A.F.A. Smit and P. Green, unpubl.; http://repeatmasker.genome.washington.edu/cgi-bin/RM2_req.pl). Because many TE insertions were consensus or mixed sequences, we only worked on localization data of TEs. For the LTR retrotransposons, we considered as insertions only the retrieved sequences with at least one complete or one incomplete typical LTR to avoid wrongly attributing very divergent copies to a given element. Each match was then checked to decide whether different copies of a given LTR retrotransposon were present, or only one copy detected by different matches. For non-LTR retrotransposons and transposons, we considered those retrieved sequences that were >400 bp, so as to eliminate short, deleted, highly divergent sequences. Each match was also checked as above.

The TE density was estimated from the number of TE insertions per base pair in the sequence fragments considered, excluding the number of N. For each class of recombination rate (see below), all genome regions corresponding to a given range of recombination rate were pooled. To analyze the relationship between TE density and recombination rate, the *Drosophila* genome was cut into nonoverlapping fragments of 0.25 Mb. The sequences corresponding to the telomeric, centromeric, and chromosome 4 regions were defined by use of the Gadfly annotations (http://hedgehog.lbl.gov:7081/

annot/). When these regions were to be excluded from the analysis, the genome was cut into nonoverlapping fragments after removing the corresponding sequences.

Estimation of Gene Densities along the Chromosomes

Because DNA sequences on each chromosome arm were not yet fully annotated, we searched for the chromosomal location of the 13376 known and predicted genes of the *Drosophila* genome by using BLAST and the data on transcribed gene sequences of release 1 (na.gadfly.dros.RELEASE1 at the Berkeley *Drosophila* Genome Project site: http://www.fruitfly.org). Gene density along the chromosomes was thus determined as described above for the TEs. To analyze the relationship between gene density and recombination rate, the *Drosophila* genome was cut into nonoverlapping fragments of 0.25 Mb.

Estimation of Recombination Rate

The rate of recombination along the chromosomes was determined by use of a procedure similar to that described by Kliman and Hey (1993). The *D. melanogaster* genetic map data was taken from FlyBase (FlyBase 1998). We selected the 892 loci that had been located in both the genetic map and the genomic sequence. The recombination rate was estimated for each chromosome arm by taking the derivative of the best fitting polynomial function of the genetic distance versus the nucleotide coordinate in the genomic sequence. Second-degree polynomial curves fitted the data set ($r2 > 0.97$) well for all chromosome arms. The fitting by the polynomial function is clearly correct for the points in the center of the curve, but deviates for the subtelomeric (chromosomal sections 20, 40–41, 80–81) and pericentromeric (chromosomal sections 1, 21, 60–61) regions, in which the recombination rates are low (Kliman and Hey 1993). To analyze the relationships between recombination rate and TE density, and between recombination rate and gene density, for each genomic fragment of 0.25 Mb (see above), the recombination rate was estimated from the value of the derivative of the polynomial curve at the middle position of the fragment. The 0.0 cM/Mb value was assigned to the fragments of the telomeric and pericentromeric regions and of chromosome 4. We also defined 49 classes of recombination rate from 0.0 to 4.9 cM/Mb at intervals of 0.1 cM/Mb. The telomeric and pericentromeric regions, and the chromosome 4 were assigned to the 0.0–0.1 cM/Mb class of recombination.

ACKNOWLEDGMENTS

We thank D. Chessel, L. Duret, and C. Vieira for their comments. This work was funded by the Centre National de la Recherche Scientifique (Programme génome, GDR 2157) and the Association pour la Recherche sur le Cancer (contract 5428 to C.B.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., Biemont, C., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185–2195.

The *Arabidopsis* genome initiative. 2000. *Nature* **408:** 796–815.

Aulard, S., Lemeunier, F., Hoogland, C., Chaminade, N., Brookfield, J.F., and Biémont, C. 1995. Chromosomal distribution and population dynamics of the *412* retrotransposon in a natural population of *Drosophila melanogaster*. *Chromosoma* **103:** 693–699.

Bachtrog, D., Weiss, S., Zangerl, B., Brem, G., and Schlötterer, C. 1999. Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Mol. Biol. Evol.* **16:** 602–610.

Baird, S.E., Sutherlin, M.E., and Emmons, S.W. 1992. Reproductive isolation in *Rhabditidae* (*Nematoda*: *Secernentea*): Mechanisms that isolate six species of three genera. *Evolution* **46:** 585–594.

Barnes, T.M., Kohara, Y., Coulson, A., and Hekimi, S. 1995. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* **141:** 159–179.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., and Rodier, F. 1985. The mosaic genome of warm-blooded vertebrates. *Science* **228:** 953–958.

Biémont, C. 1992. Population genetics of transposable DNA elements. A *Drosophila* point of view. *Genetica* **86:** 67–84.

Biémont, C. and Gautier, C. 1988. Localization polymorphism of *mdg-1*, *copia*, *I* and *P* mobile elements in genomes of *Drosophila melanogaster*, from data of inbred lines. *Heredity* **60:** 335–346.

Biémont, C., Aouar, A., and Arnault, C. 1987. Genome reshuffling of the *copia* element in a *Drosophila melanogaster* inbred line. *Nature* **329:** 742–744.

Biémont, C., Tsitrone, A., Vieira, C., and Hoogland, C. 1997. Transposable element distribution in *Drosophila*. *Genetics* **147:** 1997–1999.

Boissinot, S., Entezam, A., and Furano, A.V. 2001. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **18:** 926–935.

Boyle, A.L., Ballard, S.G., and Ward, D.C. 1990. Differential distribution of long and short interspersed element sequences in the mouse genome—Chromosome karyotyping by fluorescence *in situ* hybridization. *Proc. Natl. Acad. Sci.* **87:** 7757–7761.

Brandes A., Heslop-Harrison, J.S., Kamm, A., Kubis, S., Doubrick, R.L., and Schmidt, T. 1997. Comparative analysis of the chromosomal genomic organization of Ty1-*copia*-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol. Biol.* **33:** 11–21.

Cangiano, G. and La Volpe, A. 1993. Repetitive DNA sequences located in the terminal portion of the *Caenorhabditis elegans* chromosomes. *Nucleic Acids Res.* **21:** 1133–1139.

Charlesworth, B. 1994. Patterns in the genome. *Curr. Biol.* **4:** 182–184.

Charlesworth, D. and Charlesworth, B. 1995. Transposable elements in inbreeding and outbreeding populations. *Genetics* **140:** 415–417.

Charlesworth, B., Lapid, A., and Canada, D. 1992a. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genet. Res.* **60:** 103–114.

Charlesworth, B., Lapid, A., and Canada, D. 1992b. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. II. Inferences on the nature of selection against elements. *Genet. Res.* **60:** 115–130.

Charlesworth, D., Morgan, M.T., and Charlesworth, B. 1993. Mutation accumulation in finite outbreeding and inbreeding population. *Genet. Res.* **61:** 39–56.

Charlesworth, B., Sniegowski, P., and Stephan, W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371:** 215–220.

Charlesworth, B., Langley, C.H., and Sniegowski, P. 1997. Transposable element distributions in *Drosophila*. *Genetics* **147:** 1993–1995.

Copenhaver, G.P., Nickel, K., Kuromori, T., Benito M.I., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L.D., et al. 1999. Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286:** 2468–2474.

Costas, J., Valade, E., and Naveira, H. 2001. Structural features of the *mdg1* lineage of the *Ty3/gypsy* group of LTR retrotransposons inferred from the phylogenetic analyses of its open reading frames. *J. Mol. Evol.* **53:** 165–171.

Duret, L., Marais, G., and Biémont, C. 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. *Genetics* **156:** 1661–1669.

Eanes, W.F., Wesley, C., and Charlesworth, B. 1992. Accumulation of P elements in minority inversions in natural populations of *Drosophila melanogaster*. *Genet. Res.* **59:** 1–9.

Eissenberg, J.C. and Hilliker, A.J. 2000. Shedding new light on the dark corners of the nucleus. *Genetica* **109:** 3–6.

Finnegan, D.J. 1992. Transposable elements. In The genome of *Drosophila melanogaster*. (ed. D.L. Lindsley, and G. Zimm,) pp. 1096–1107, Academic Press, New York.

Fullerton, S.M., Carvalho, A.B., and Clark, A.G. 2001. Local rates of recombination are positively correlated with GC content in the human genome. *Mol. Biol. Evol.* **18:** 1139–1142.

Garcia Guerreiro, M.P. and Fontdevila, A. 2001. Chromosomal distribution of the transposable elements *Osvaldo* and *blanco* in

original and colonizer populations of *Drosophila buzzatii*. *Genet. Res.* **77:** 227–238.
Goldman, A.S.H. and Lichten, M. 1996. The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. *Genetics* **144:** 43–55.
———. 2000. Restriction of ectopic recombination by interhomolog interactions during *Saccharomyces cerevisiae* meiosis. *Proc. Natl. Acad. Sci.* **97:** 9537–9542.
Gray, Y.H.M. 2000. It takes two transposons to tango. Transposable-elements-mediated chromosomal rearrangements. *Trends Genet.* **16:** 461–468.
Heslop-Harrison J.S., Brandes, A., Taketa, S., Schmidt T., Vershinin, A.V., Alkhimova, E.G., Kamm, A., Doudrick, R.L., Schwarzacher, T., Katsiotis, A., et al. 1997. The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* **100:** 197–204.
Hill, W.G. and Robertson, A. 1966. The effect of linkage on the limit to artificial selection. *Genet. Res.* **8:** 269–294.
Hoogland, C. and Biémont, C. 1996. Chromosomal distribution of transposable elements in *Drosophila melanogaster*. Test of the ectopic recombination model for maintenance of insertion site number. *Genetics* **144:** 197–204.
Jabbari, K. and Bernardi, G. 2000. The distribution of genes in the *Drosophila* genome. *Gene* **247:** 287–292.
Jarne, P., David, P., and Viard, F. 1998. Microsatellites, transposable elements and the X chromosome. *Mol. Biol. Evol.* **15:** 28–34.
Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci.* **94:** 1872–1877.
Kalendar, R., Grob, T., Regina, M., Suoniemi, A., and Shulman, A.H. 1999. IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* **98:** 704–711.
Kidwell, M.G. and Lisch, D.R. 1997. Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci.* **94:** 7704–7711.
Kidwell, M.G. and Evgen'ev, M.B. 1999. How valuable are model organisms for transposable element studies? *Genetica* **107:** 103–111.
Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A., and Voytas, D.F. 1998. Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.* **8:** 464–478.
Kleckner, N. and Weiner, B. 1993. Potential advantages of unstable interactions for pairing of chromosomes in meiotic, somatic and premeiotic cells. *Cold Spring Harbor Symp. Quant. Biol.* **58:** 553–565.
Kliman, R.M. and Hey, J. 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol. Biol. Evol.* **10:** 1239–1258.
Korenberg, J.R. and Rykowski, M.C. 1988. Human genome organization: *Alu*, LINEs, and the molecular structure of metaphase chromosome bands. *Cell* **53:** 391–400.
Kumar, A. and Bennetzen J.L. 1999. Plant retrotransposons. *Annu. Rev. Genet.* **33:** 479–532.
Kupiec, M. and Petes, T.D. 1988. Meiotic recombination between repeated transposable elements in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8:** 2942–2954.
Langley, C.H., Montgomery, E.A., Hudson, R., Kaplan, N., and Charlesworth, B. 1988. On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res.* **52:** 223–235.
Le, Q.H., Wright, S., Yu, Z., and Bureau, T. 2000. Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **97:** 7376–7381.
Lerat, E., Biémont, C., and Capy, C. 2000. Codon usage and the origin of *P* elements. *Mol. Biol. Evol.* **17:** 467–468.
Lerat, E., Capy, P., and Biémont, C. 2002. Codon usage by transposable elements and their host genes in five species. *J. Mol. Evol.* (in press).
Liao, G.C., Rehm, E.J., and Rubin, G.M. 2000. Insertion site preferences of the *P* transposable element in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **97:** 3347–3351.
Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I., Town, C.D., Fujii, C.Y., Mason, T., Bowman, C.L., Barnstead, M., et al., 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402:** 761–768.
Marais, G., Mouchiroud, D., and Duret, L. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl. Acad. Sci.* **98:** 5688–5692.
McKim, K.S., Green-Marroquin, B.L., Sekelsky, J.J., Chin, G., Steinberg, C., Khodosh, R., and Hawley, R.S. 1998. Meiotic synapsis in the absence of recombination. *Science* **279:** 876–878.
Montgomery, E.A., Charlesworth, B., and Langley, C.H. 1987. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet. Res.* **49:** 31–41.
Montgomery, E.A., Huang, S.M., Langley, C.H., and Judd, B.H. 1991. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: Genome structure and evolution. *Genetics* **129:** 1085–1098.
Morgan, M.T. 2001. Transposable element number in mixed mating populations. *Genet. Res.* **77:** 261–275.
Nadir, E., Margalit, H., Gallily, T., and Ben-Sasson, S.A. 1996. Microsatellite spreading in the human genome: Evolutionary mechanisms and structural implications. *Proc. Natl. Acad. Sci.* **93:** 6470–6475.
Nuzhdin, S.V. 1999. Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica* **107:** 129–137.
Pardue, M.L., Danilevskaya, O.N., Lowenhaupt, K., Slot, F., and Traverse, K.L. 1996. *Drosophila* telomeres: New views on chromosome evolution. *Trends Genet.* **12:** 48–52.
Petes, T.D. and Hill, C.W. 1988. Recombination between repeated genes in microorganisms. *Annu. Rev. Genet.* **22:** 147–168.
Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., Caizzi, R., Caggnese, C., and Gatti, M. 1995. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proc. Natl. Acad. Sci.* **92:** 3804–3808.
Roy, A.M., Carroll, M.L., Kass, D.H., Nguyen, S.V., Salem, A.H., Batzer, M.A., and Deininger, P.L. 1999. Recently integrated human Alu repeats: Finding needles in the haystack. *Genetica* **107:** 149–161.
Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109:** 365–371.
Shapiro, J.A. 1999. Transposable elements as the key to a 21st century view of evolution. *Genetica* **107:** 171–179.
Sharp, P.M. and Matassi, G. 1994. Codon usage and genome evolution. *Curr. Opin. Genet. Dev.* **4:** 851–860.
Smit, A.F.A. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9:** 657–663.
Smit, A.F.A. and Green, P. 2000. RepeatMasker. (http://repeatmasker.genome.washington.edu/cgi-bin/RM2_req.pl).
Sniegowski, P. and Charlesworth, B. 1994. Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. *Genetics* **137:** 815–827.
Springer, M.S., Tusneem, N.A., Davidson, E.H., and Britten, R.J. 1995. Phylogeny, rates of evolution, and patterns of codon usage among sea urchin retroviral-like elements, with implications for the recognition of horizontal transfer. *Mol. Biol. Evol.* **12:** 219–230.
Sun, X.P., Wahlstrom, J., and Karpen, G. 1997. Molecular structure of a functional *Drosophila* centromere. *Cell* **91:** 1007–1019.
Tomilin, N.V. 1999. Control of genes by mammalian retroposons. *Int. Rev. Cytol.* **186:** 1–48.
Toth, G., Gaspari, Z., and Jurka, J. 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* **10:** 967–981.
Virgin, J.B. and Bailey, J.P. 1998. The M26 hotspot of *Schizosaccharomyces pombe* stimulates meiotic ectopic recombination and chromosomal rearrangements. *Genetics* **149:** 1191–1204.
Wright, S.I. and Schoen, D.J. 1999. Transposon dynamics and the breeding system. *Genetica* **107:** 139–148.
Zhang, J. and Peterson, T. 1999. Genome rearrangements by nonlinear transposons in maize. *Genetics* **153:** 1403–1410.
Zuckerkandl, E. and Hennig, W. 1995. Tracking heterochromatin. *Chromosoma* **104:** 75–83.

WEB SITE REFERENCES

http: //flybase.bio.indiana.edu/; the FlyBase database of the *Drosophila* Genome Projects and community literature. FlyBase 1998.
http: //hedgehog.lbl.gov:7081/annot/; the genome annotation database. GadFly 2000.
http: //repeatmasker.genome.washington.edu/cgi/RM2_req.pl; `RepeatMasker` Smit, A.F.A. and Green, P. 2000.
http: //www.ebi.ac.uk/index.html; the European Bioinformatics Institute 2001.
http: //www.fruitfly.org/index.html; the Berkeley *Drosophila* Genome Project 2000.

Received August 15, 2001; accepted in revised form December 14, 2001.